PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-305832

(43)Date of publication of application : 02.11.2000

(51)Int.Cl. G06F 12/00
G06F 9/46
G06F 15/16

(21)Application number : 11-116853

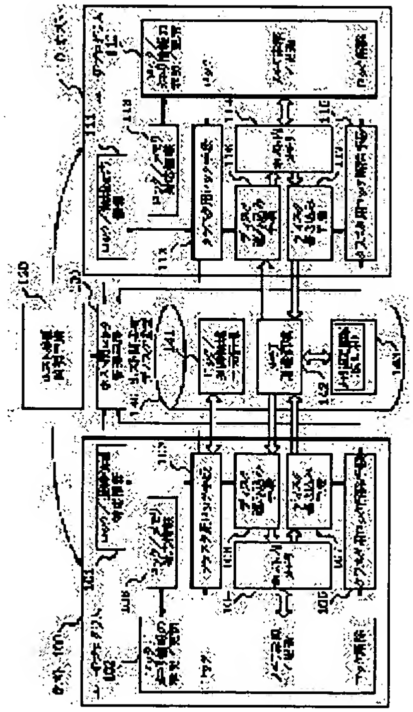
(22)Date of filing : 23.04.1999

(71)Applicant : NEC SOFTWARE KYUSHU LTD

(72)Inventor : OTSUKA HIDEAKI

(54) DEVICE AND METHOD FOR SHARING MEMORY IN CLUSTER SYSTEM CONSTITUTED OF PLURAL HOSTS

(57)Abstract:



PROBLEM TO BE SOLVED: To transfer many user programs operated on a system constructed by one host to a cluster system constituted of plural hosts without modifying these programs.

SOLUTION: A host 100 includes a user program 102, a cluster locking means 103, an intra-host memory 104, a cluster unlocking means 105, a disk reading means 106, a disk writing means 107, lock/memory correspondence information 108, and lock/retreat area correspondence information 101 and a host 110 also has similar functions. A host operation monitoring means 120, an inter-host lock management means 130 and an inter-host sharing disk device 140 are prepared as shared functions between plural hosts. Since the contents of the intra-host memory 104 are automatically updated at the time of executing the locking means 103 and the unlocking means 105, user programs sharing the memory can be transferred from plural processes driven in one host to the cluster system constituted of plural hosts and executed without modifying these programs.

LEGAL STATUS

[Date of request for examination] 24.03.2000

[Date of sending the examiner's decision of rejection] 13.04.2004

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the share equipment and the approach of memor in the cluster system constituted from two or more processes which have a track record of operation especially within 1 host by two or more hosts who enable activation of the user progra which shares memory with the cluster system constituted without modification of a user program by two or more hosts about the share equipment and the approach of memory in the cluster syste which consists of two or more hosts.
[0002]
[Description of the Prior Art] The memory share method in the case of the multi-process system of the conventional memory share mold is indicated by JP,08-166933,A. This conventional memo share method consists of a host share data area which exists in each host, a host share data area whereabouts management tool, and a data transceiver means.
[0003] The conventional memory share method which has such a configuration operates as follows.
[0004] That is, if a program accesses a host share data area, the host share data area manageme tool would operate, it would communicate with other hosts, and the host share data area will be amended.
[0005]
[Problem(s) to be Solved by the Invention] I hear that the 1st trouble needs reconstruction of the existing user program which had already operated on 1 host, and the conventional memory share method mentioned above has it.
[0006] The reason is for having to convert so that a host share data area may be accessed.
[0007] I hear that the 2nd trouble has the late processing engine performance, and it has it.
[0008] The reason needs the communication link with other hosts, and is for traffic to increase i proportion to the number of a host. Moreover, it is the point that the check to other hosts is need to access memory exclusively.
[0009]
[Means for Solving the Problem] The share equipment and the approach of memory in the cluste system which consists of two or more hosts of this invention The user program which shares memory from two or more processes, and the lock means for clusters, The memory in a host, the lock discharge means for clusters, and a disk reading means, With the host having a disk write-in means, the information corresponding to a lock/memory, and the information corresponding to a lock/save area Among two or more hosts, a host operation monitor means and the lock management means between hosts, Have shared disk equipment between hosts and each host's operation situation is supervised with a host operation monitor means. The lock management means between hosts performs lock control among hosts, and the shared disk equipment between hosts is constituted from a lock identifier, the lock / save area unitary information that a memory save area is managed, a memory save area that exists for every lock identifier, and a memory sa

area (for restoration). Information is managed by the key in a lock identifier with the memory address of the memory in a host whose information corresponding to the lock/memory which exis for every host is a shared memory corresponding to a lock identifier and its lock identifier, and th memory size of the memory in a host. The lock / save area unitary information recorded on the information corresponding to the lock/save area which exists for every host, and the shared disk equipment between hosts the information on the memory save area which is a save area of the shared memory corresponding to a lock identifier and its lock identifier Information is managed b the key in a lock identifier with the save area information (for restoration) which manages the sa area information to manage and the information on a memory save area (for restoration). The memory save area of the shared disk equipment between hosts during the record writing to the shared disk equipment between hosts by host down In CHIEEKU of the justification of a record when an access failure occurs, if the data of the write-in check field A of the head section of a record and the write-in check field B of a trailer are right, the value of a memory save area in th meantime is also constituted so that the record format a right thing is guaranteed to be may perform. When a user program is first performed within a host, the information corresponding to lock/memory It is created synchronizing with partitioning of the memory in a host, or acquisition a lock identifier. The information corresponding to a lock/save area If the information corresponding to the lock identifier as which it was referred to at the time of the lock means activation for clusters, and the lock was required from the user program is not registered If not registered with reference to a lock / save area unitary information, a memory save area and a memory save area (for restoration) are newly secured to the lock identifiers. The information for accessing the field is registered into a lock / save area unitary information, and the information corresponding to a lock/save area. A memory save area and a memory save area (for restoration are initialized at the time of reservation of a field. The lock means for clusters To the lock management means between hosts, require a lock and a disk reading means is used after a lock success with reference to the information and the information corresponding to a lock/memory corresponding to a lock/save area. Data are read into the memory in a host from a memory save area, and a user program performs the reference and updating of the memory in a host after a loc The lock discharge means for clusters A disk write-in means is used with reference to the information and the information corresponding to a lock/memory corresponding to a lock/save ar After writing the data of the memory in a host in a memory save area, lock discharge is required from the lock management means between hosts. The lock management means between hosts While managing the lock demand from each host, when a certain host is downed, By canceling the lock whose downed host received the notice and acquired it from the host operation monitor means, and updating the memory in a host automatically at the time of the lock means for cluster and the lock discharge means activation for clusters It is constituted so that activation of the use program which shares memory from two or more processes which were operating within the hos may be enabled with the cluster system constituted without modification of a user program by tw or more hosts.

[0010]

[Embodiment of the Invention] This invention is a method for enabling activation of the user program which shares memory from two or more processes with a track record of operation with 1 host with the cluster system constituted without modification of a user program by two or more hosts. The cluster system said here is a system which mounts the system of the same configuration in two or more sets of hosts, and carries out the load distribution of the processing automatically according to each host's load profile initiation.

[0011] Next, the gestalt of operation of this invention is explained with reference to a drawing.

[0012] Drawing 1 is the block diagram showing the share equipment of memory and the gestalt o 1 operation of an approach in the cluster system which consists of two or more hosts of this invention.

[0013] Reference of drawing 1 constitutes the gestalt of operation of this invention from shared disk equipment 140 between hosts shared with a host 100, a host 110, a host operation monitor means 120 to supervise operation of each host, and a lock management means 130 between host to perform lock control between hosts, among hosts.

[0014] To the host 100 who constitutes a cluster system A user program 102 and a lock means 103 for clusters to offer the user interface of a lock function, The memory 104 in a host which is the memory area which a user program 102 shares with other programs, A lock discharge means 105 for clusters to offer the user interface of a lock discharge function, A disk reading means 10 to read data into the memory 104 in a host from the shared disk equipment 140 between hosts, T disk write-in means 107 which writes data in the shared disk equipment 140 between hosts from the memory 104 in a host, It consists of a lock identifier, information 108 corresponding to a lock/memory that correspondence of a shared memory field is managed, and information 101 corresponding to a lock/save area that correspondence of the save area of a lock identifier and memory is managed.

[0015] A host 110 can increase the number of a host freely, although it is the same configuration as a host 100 and has two hosts' composition in drawing 1 .

[0016] The host operation monitor means 120 is used in order to supervise each host's operation situation.

[0017] The lock management means 130 between hosts is used in order to perform lock control among hosts.

[0018] The shared disk equipment 140 between hosts consists of a lock / save area unitary information 141 that a lock identifier and a memory save area are managed, and the memory save area 142 which exists for every lock identifier and the memory save area (for restoration) 143.

[0019] To the information 108 corresponding to the lock/memory which exists for every host Th share equipment of memory in the cluster system which consists of two or more hosts of this invention of drawing 3 , and the lock identifier in an approach, As shown in drawing showing the memory address of the memory in a host which is a shared memory corresponding to the lock identifier, and the structure of the correspondence information on memory size It has the lock identifier 301, the memory address 302 of the memory 104 in a host which is a shared memory corresponding to the lock identifier, and the memory size 303 of the memory 104 in a host, and information is managed by the key in the lock identifier 301.

[0020] To the lock / save area unitary information 141 recorded on the information 101 corresponding to the lock/save area which exists for every host, and the shared disk equipment 140 between hosts The share equipment of memory in the cluster system which consists of two o more hosts of this invention of drawing 4 R> 4, and the lock identifier in an approach, As shown i drawing showing the structure of the correspondence information on the save area information o the shared memory corresponding to the lock identifier, and save area information (for restoratio The lock identifier 401 and the save area information 402 which manages the information on the memory save area 142 which is a save area of the shared memory corresponding to the lock identifier, It has the save area information (for restoration) 403 which manages the information o the memory save area (for restoration) 143, and information is managed by the key in the lock identifier 401.

[0021] The memory save area 142 of the shared disk equipment 140 between hosts consists of a write-in check field A501, a memory save area 502, and a write-in check field B503, as shown in drawing showing the structure of the correspondence information on the write-in check field of t memory save area of the share equipment of memory in the cluster system which consists of two or more hosts of this invention of drawing 5 , and the shared disk equipment between hosts in an approach, a memory save area, and a write-in check field.

[0022] These means operate as follows, respectively. In addition, although explanation is given about a means to share between hosts 100 etc., each means of a host 110 also carries out the sa actuation.

[0023] The information 108 corresponding to a lock/memory is created synchronizing with partitioning of the memory 104 in a host, or acquisition of a lock identifier, when a user program 102 is first performed within a host 100.

[0024] If the information corresponding to the lock identifier as which the information 101 corresponding to a lock/save area was referred to at the time of the lock means 103 activation fo clusters, and the lock was required from the user program 102 is not registered If not registered there with reference to a lock / save area unitary information 141 The memory save area 142 an the memory save area (for restoration) 143 are newly secured to the lock identifiers, and the information for accessing the field is registered into a lock / save area unitary information 141, a the information 101 corresponding to a lock/save area.

[0025] The memory save area 142 and the memory save area (for restoration) 143 are initialized at the time of reservation of a field.

[0026] The lock means 103 for clusters To the cluster system which consists of two or more ho of this invention of drawing 2 The host 200 top who is the memory share mold multiprocessor system shown in the block diagram showing the multi-process system of the memory share mold which is one host who constitutes the cluster system in the share equipment and the approach of memory which can be set, Have the user interface which is compatible with the lock means 205, and the lock management means 130 between hosts is received. A lock is required and data are read into the memory 104 in a host from the memory save area 142 after a lock success using th disk reading means 106 with reference to the information 101 and the information 108 corresponding to a lock/memory corresponding to a lock/save area.

[0027] A user program 102 performs the reference and updating of the memory 104 in a host aft a lock.

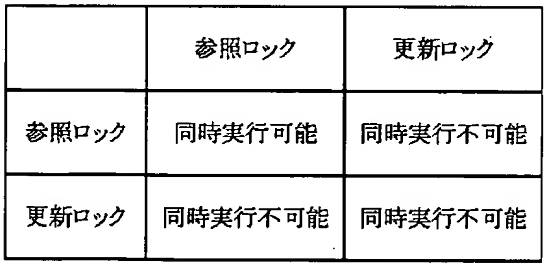
[0028] A user program 102 is the same program as the user program 202 and user program 203 which operate on the multiprocessor system of the memory share mold shown by drawing 2 .

[0029] After the lock discharge means 105 for clusters has a user interface with the lock discharge means 207 and the compatibility on the host 200 who is the memory share mold multiprocessor system shown by drawing 2 and writing the data of the memory 104 in a host in a memory save area 142 using the disk write-in means 107 with reference to the information 101 and the information 108 corresponding to a lock/memory corresponding to a lock/save area, lock discharge requires from the lock management means 130 between hosts.

[0030] The lock management means 130 between hosts manages the lock demand from each hos and offers service as shown in the example of lock control of Table 1.

[0031]

[Table 1]



	参照ロック	更新ロック
参照ロック	同時実行可能	同時実行不可能
更新ロック	同時実行不可能	同時実行不可能

[0032] Moreover, the lock management means 130 between hosts cancels the lock whose downe host received the notice and acquired it from the host operation monitor means 120, when a certa host is downed.

[0033] As a property of the shared disk equipment 140 between hosts, when an access failure occurs by host down etc. during record writing, if the data of the head section of a record and a trailer are right, the value in the meantime also shows the record format in case a right thing is guaranteed to drawing 5 . It is necessary to change into the record format which suited the property of the actually used disk unit.

[0034] When the data writing to the memory save area 142 goes wrong by host down etc., the memory save area (for restoration) 143 is used in order to restore the memory save area 142.

[0035] Next, actuation of the gestalt of operation of this invention is explained to a detail with reference to drawing 1 - drawing 5 , and Table 1.
[0036] actuation is explained conventionally on the multi-process system of the memory share mold which is one host who constitutes a cluster system by measure introduction and drawing 2 .
[0037] The user programs 202 or 203 which operate by the host 200 perform partitioning of the memory 206 in a host, and acquisition of a lock identifier, when it performs first within a host, an they register them into the information 204 corresponding to a lock/memory.
[0038] The information 204 corresponding to a lock/memory is referred to from user programs 202 or 203, and is used for acquisition of the lock at the time of memory 206 access in a host, a memory address, etc.
[0039] If a lock demand is performed from user programs 202 or 203, the lock means 205 will lo using the lock management means 201 in a host. The lock management means 201 in a host offer service as shown in Table 1.
[0040] User programs 202 or 203 access memory 206 in a host after a lock.
[0041] If a lock discharge demand is performed from user programs 202 or 203, the lock dischar means 207 will perform lock discharge using the lock management means 201 in a host.
[0042] Next, the actuation in the case of a cluster system is explained.
[0043] A user program 102 performs partitioning of the memory 108 in a host, and acquisition of lock identifier, when it performs first within a host 100, and it registers them into the information 108 corresponding to a lock/memory.
[0044] In order to perform a user program 102 and to lock the memory 104 in a host at the time reference or updating, the lock means 103 for clusters is performed. The lock means 103 for clusters performs the next actuation with the mode of the lock which the user program 102 required, when it cooperates with the lock management means 130 between hosts and the lock o the whole cluster system is successful.
[0045] (11) In a reference lock, since it is the purpose to refer to a shared memory exclusively, reference lock operates as follows.
[0046] With reference to the information 108 and the information 101 corresponding to a lock/sa area corresponding to a lock/memory, the lock means 103 for clusters acquires the memory address 302 corresponding to the lock identifier shown by drawing 3 and drawing 4 and memory size 303, the save area information 402, and the save area information (for restoration) 403 from user program 102 to a lock demand, and reads data into the memory 104 in a host from the memory save area 142 of the shared disk equipment 140 between hosts at it using the disk readi means 106. At this time, after checking the check field B503, writing in with the write-in check field A501 shown in drawing 5 in the memory save area 142, reading data into the memory 104 in host from the memory save area (for restoration) 143 by host down etc. when data are inaccurate and writing the information on the memory save area (for restoration) 143 in the memory save ar 142, control is returned to a user program 102.
[0047] The lock discharge means 105 for clusters cooperates with the lock management means 130 between hosts to the lock discharge demand of a user program 102 to a reference lock, and performs lock discharge to it.
[0048] (12) In the case of an update lock, since it is the purpose to update a shared memory exclusively, an update lock operates as follows.
[0049] With reference to the information 108 and the information 101 corresponding to a lock/sa area corresponding to a lock/memory, the lock means 103 for clusters acquires the memory address 302 corresponding to the lock identifier shown by drawing 3 and drawing 4 and memory size 303, the save area information 402, and the save area information (for restoration) 403 from user program 102 to a lock demand, and reads data into the memory 104 in a host from the memory save area 142 of the shared disk equipment 140 between hosts at it using the disk readi means 106. At this time, after checking the check field B503, writing in with the write-in check field A501 shown in drawing 5 in the memory save area 142, reading data into the memory 104 in

host from the memory save area (for restoration) 143 by host down etc. when data are inaccurate and writing the information on the memory save area (for restoration) 143 in the memory save ar 142, control is returned to a user program 102. In a right case, the data of the memory save area 142 write the contents of the memory save area 142 in the memory save area (for restoration) 1 using the disk write-in means 107.

[0050] The lock discharge means 105 for clusters writes the memory 104 in a host in the lock discharge demand of an update lock from a user program 102 in the memory save area 142 of the shared disk equipment 140 between hosts using the disk write-in means 107, it cooperates with the lock management means 130 between hosts, and lock discharge is performed.

[0051] Next, actuation when a host down occurs during a lock in a certain host is explained.

[0052] (21) In a reference lock, since it is the purpose that a reference lock refers to a shared memory exclusively, the restoration processing to memory is unnecessary.

[0053] About a lock, the case where a host down occurs in a host 110 is explained to an example The host operation monitor means 120 will be notified to the lock management means 130 betwe hosts, if a host's 110 down is recognized. The lock management means 130 between hosts cance all the locks from a host 110. The user program of the host 100 whose actuation was attained by lock discharge performs the above-mentioned processing.

[0054] (22) In the case of an update lock, although it is the purpose that an update lock updates a shared memory exclusively, the restoration processing to memory is unnecessary.

[0055] About discharge of a lock, it operates by the same structure as the time of a reference lock. The user program of the host 100 whose actuation was attained by lock discharge performs the above-mentioned processing. Even when the data of the memory 104 in a host are written in the memory save area 142 of the shared disk equipment 140 between hosts and a failure occurs inside, since the check of data is performed at the time of a lock, the right data at the time of the last lock discharge success can be referred to.

[0056] Next, it explains using a concrete example.

[0057] (31) When both a user program 102 and the user program 112 are reference locks, a hos 100 user program 102 performs the lock (reference lock) for referring to a shared memory with reference to the information 108 corresponding to lock/. At this time, data are read into the memory 104 in a host from the memory save area 142 in the shared disk equipment 140 between hosts. Next, a lock (reference lock) for a host's 110 user program 112 to refer to a shared memo with reference to the information 118 corresponding to lock/is performed. Since it is a reference lock, it succeeds in a lock, and data are read into the memory 104 in a host like a user program 102.

[0058] (32) With a reference lock, a user program 102 performs a lock (reference lock) for a host's 100 user program 102 to refer to a shared memory with reference to the information 108 corresponding to lock/, when a user program 112 is an update lock. At this time, data are read in the memory 104 in a host from the memory save area 142 in the shared disk equipment 140 between hosts. Next, since it is [reference] under lock by the user program 102 when a lock (update lock) for a host's 110 user program 112 to update a shared memory with reference to the information 118 corresponding to lock/is performed, it will be in the state waiting for a lock. Afte lock is canceled by the user program 102, a lock is successful, and data are read into the memor 104 in a host from the memory save area 142 in the shared disk equipment 140 between hosts. The memory 114 in a host is written in the memory save area 142 in the shared disk equipment 140 between hosts by the user program 112 at the time of lock discharge.

[0059] (33) When both a user program 102 and the user program 112 are update locks, a host's 100 user program 102 performs the lock (update lock) for updating a shared memory with reference to the information 108 corresponding to lock/. At this time, data are read into the memory 104 in a host from the memory save area 142 in the shared disk equipment 140 between hosts. Next, since it is among an update lock in a user program 102 when a lock (update lock) for host's 110 user program 112 to update a shared memory with reference to the information 118 corresponding to lock/is performed, it will be in the state waiting for a lock. A user program 102 written in the memory save area 142 in the shared disk equipment 140 between hosts at the time of lock discharge, after a lock is canceled, as for the memory 104 in a host, a lock is successful b the user program 112, and data are read into the memory 104 in a host from the memory save ar 142 in the shared disk equipment 140 between hosts. The memory 114 in a host is written in the memory save area 142 in the shared disk equipment 140 between hosts by the user program 112 the time of lock discharge.

[0060]

[Effect of the Invention] As explained above, the share equipment and the approach of memory i the cluster system which consists of two or more hosts of this invention have the 1st effectivene in the ability to shift to the cluster system which consists of two or more hosts without reconstruction of the user program of a large number which are operating on the system which w being built by one host.

[0061] Although the reason can perform a user program only within 1 host when the memory sha interlocked with lock/unlock facility cannot be performed, it is because a user program can be performed without reconstruction by two or more hosts by using this invention. Moreover, althou there is also a method of not using this invention but converting a user program into the processi logic suitable for a cluster system, it is **** about a great correction man day in that case.

[Translation done.]